Breast image pre-processing for mammographic tissue segmentation

Image Processing and Computer Vision's Final Project by David Morales and Anastasia Kuflievskaya

***Abstract*—The focus of this project is to develop and implement a pre-processing technique for mammographic images, so as to enhance tissue segmentation of a mammogram. In view of the methods that have been followed in some relevant literature, we have applied a sequence of considered processing steps that includes periphery separation, intensity ratio propagation, breast thickness estimation, and intensity balancing. These techniques address common issues like uneven illumination and intensity variations that may hamper accurate image analysis. Our results indicate a better breast tissue segmentation and visualization, therefore enabling more accurate breast cancer diagnosis.**

***Index Terms*—Breast thickness estimation, breast tissue segmentation, intensity Balancing, mammographic image processing, Otsu's thresholding, peripheral enhancement.**

I. Introduction

BREAST cancer is one of the most prevalent cancers among women across the world and it represents the second leading cause of cancer death among them [1], making its detection and diagnosis at an early stage of the utmost importance. Indeed, its early detection can improve the chances of successful treatment and reduce its mortality rate [2]. With this purpose, mammography is a powerful tool, involving low-energy X-rays. It's extremely helpful in detecting any irregularity in the breast tissue.

However, the quality of the images varies considerably. There are difficulties with nonuniform illumination and inconsistency in intensity. To handle these issues, in this paper, we have implemented a sequence of well-known image pre-processing techniques.

This paper aims to follow the approach in [3] for the enhancement of the breast tissue, making it more distinguishable. The steps that have been considered are the following: breast periphery separation from the background, correction of intensity variations by means of intensity ratio propagation, breast thickness estimation, and finally, intensity balancing for uniform illumination. These improvements are highly relevant to obtain accurate segmentation of breast tissue, which is essential in clinical practice for the assessment of accurate breast cancer diagnosis.

In order to apply the method described above, two images have been extracted from a dataset [4]: a craniocaudal mammographic image (CC) and a mediolateral oblique mammographic image (MLO).

II. Breast Periphery Separation

The separation of the breast periphery makes the breast tissue more distinguishable from the background. At its core, it involves the isolation of the breast area from irrelevant parts of the image. The removal of these components provides a clearer and more focused view of the breast tissue itself. This is therefore really important for the following tasks of image analysis. A well-defined breast periphery ensures that the algorithms used for further processing can work more effectively.

A. Method

Otsu's thresholding: Otsu's technique [5] is used to find the optimal threshold that differentiates the breast tissue from the background. The resulting thresholded image highlights regions where intensity values exceed this threshold.

Mean intensity thresholding: In the Otsu's thresholded image, the average intensity of the breast tissue is computed and then used as a second threshold in a subsequent step.

Combining thresholds: The outcomes of Otsu's thresholding and the mean intensity thresholding are merged using a logical OR operation. This combined approach enhances the accuracy of separating the breast tissue from the background.

Hole filling: Small gaps in the binary image are filled using morphological operations [6].

Dilation: The binary image is dilated to include boundary pixels, ensuring the entire breast region is captured.

Region labeling and selection: The connected areas in the dilated binary image are labeled [7]. The largest area, representing the breast tissue, is identified and isolated.

Contour detection: The periphery of the breast is detected, which outlines the boundary of the breast tissue.

B. Visualization of results

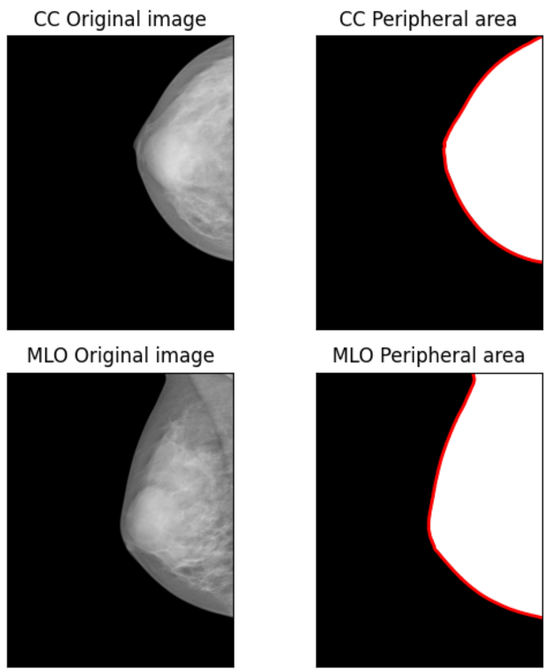

Fig. 1. The images on the left depict the original images, while the images on the right depict the isolated breast and the separated breast periphery, highlighted in red.

III. Intensity ratio propagation

The propagation of intensity ratios is a step that corrects all the intensity variations across the mammographic image. It standardizes illumination levels so that structures within the breast tissue become more visible, correcting both under- and over-illuminated areas of the image. The result is a more uniform image that makes more prominent the differences in the structure of the breast tissue.

A. Method

As the approach described in the paper [3] produced minimal changes in the image, a slightly different methodology has been carried out. The procedure used is described below.

Local neighborhood definition: A local neighborhood is defined around each pixel, capturing local intensity variations.

Intensity ratio calculation: For every pixel within the image, the average intensity of its local neighborhood is computed. This average is then used to determine an intensity correction ratio.

Intensity adjustment: Each pixel's intensity is adjusted by multiplying it with the calculated local intensity ratio. This process ensures that pixel intensities are standardized across the image.

B. Visualization of results

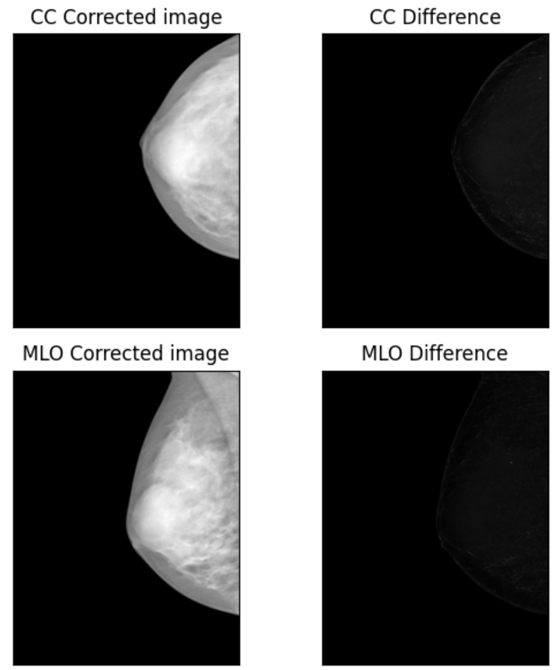

Fig. 2. Intensity ratio propagation in CC and MLO images. Both images depict an improved uniformity in tissue illumination, compared to the original images in Fig. 1. Left: corrected images using intensity ratio propagation. Right: the absolute difference between the original images and the corrected ones.

IV. Breast thickness estimation

To account for the intensity variation caused by differences in breast thickness and tissue composition, the relative breast thickness ratios are calculated. This is done by using the following formula:

$$R = \frac{p\text{Line}(P)}{p\text{Line}(P_{\text{ref}})}, \tag{1}$$

where $p\text{Line}(P)$ represents the length of the line at a specific point P, and P_{ref} is a reference point, selected as the thickest point within the breast. These ratios are also called ”length ratios”.

A. Method

Contour analysis: The separated breast periphery of the breast is analyzed to locate the farthest point from the chest wall, which is generally located near the nipple.

Contour segmentation: This contour is then divided into upper and lower sections, using the farthest point as a reference.

Construction of parallel lines: Reference points close to the top and right edges of the breast contour are found and an initial line is drawn between them. Subsequent parallel lines are then generated based on the slope of this initial line, connecting each point of the upper contour with a corresponding point of the lower contour.

Identification of the thickest point: The thickest point is determined by adding a small distance in the x-axis to the farthest point from the chest wall. Then, using the previously determined slope, a line intercepting this new point is drawn. The length of this line is known as $p\mathrm{Line}(P_{\mathrm{ref}})$.

Length ratios calculation: Length ratios are obtained by measuring the length of all parallel lines drawn between the upper and lower parts of the breast contour. The ratio for each line is found with respect to the length of the reference line, by using (1). These ratios represent the relative thickness of the breast tissue at different points, which is one of the most important parameters for understanding intensity variations across the image.

Length ratios propagation: These length ratios are then propagated across the whole image, which correct the intensity variations due to the differences in breast thickness and tissue composition. This propagation is based on the distances from each pixel to the nearest point of the breast contour. This step gives all pixels a more uniform representation of the breast tissue. It ensures that the intensity corrections get distributed smoothly over the whole image.

B. Visualization of results

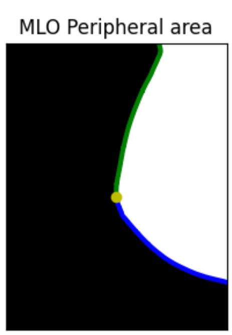

Fig. 3. Breast periphery contour. In yellow, the furthest point from the chest wall is identified. The upper and lower contours are drawn in green and in blue, respectively.

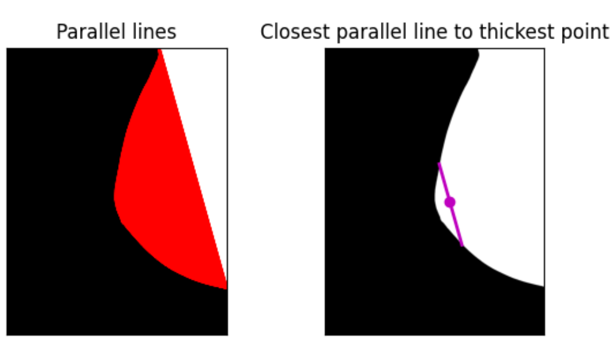

Fig. 4. Left: parallel lines between every point in the upper and lower contours. Right: thickest point and its closest parallel line, used for breast thickness estimation.

V. Intensity balancing

The aim of the intensity balancing method is to uniform the brightness of a given image. Here, pixel intensities are modified according to its distance to the skinline. This technique applies the previously propagated length ratios in order to balance the intensity in a systematic way. This adjustment ensures that the entire image is covered with a uniform distribution of illumination according to the specified length ratios and the spatial arrangement of the skinline.

A. Method

Logarithm of length ratios: Assuming a non-linear relationship between the tissue thickness and log-exposure, the logarithm of the length ratios $R(x, y)$ is computed.

Normalization of length ratios: The normalized ratios are calculated according to:

$$RP(x, y) = \frac{R(x, y) - R_{\mathrm{min}}}{R_{\mathrm{max}} - R_{\mathrm{min}}}, \tag{2}$$

where $RP(x, y)$ is the normalized ratio at pixel (x, y) and R_{min}, R_{max} are the minimum and maximum ratios.

Definition of a global thickness reference: The global thickness reference R_{ref} is defined as the mean of these normalized ratios. The normalized reference ratio is also calculated (RP_{ref}).

Distance transform: A binary mask of the skinline is generated, and the distance from each pixel to the closest skinline point is calculated.

Intensity adjustment: For every pixel within the image, the distance calculated above is used to determine the corresponding ratio index. Then, this index is used to compute the adjusted pixel intensity $P(x, y)$, given by:

$$P''(x, y) = P'(x, y)(1 + (RP_{ref} - RP(x, y))), \tag{3}$$

where $P'(x, y)$ is the original pixel intensity.

B. Visualization of results

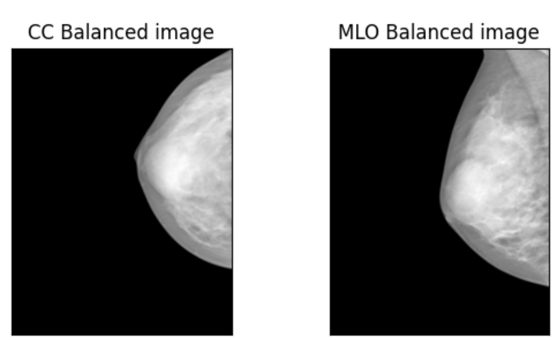

Fig. 5. Intensity balancing in CC and MLO mammographic images. Both depict a more uniform brightness and illumnination than the original images in Fig. 1.

VI. Segmentation

Segmentation is the process of delineating different tissue types within the mammographic image. However, due to a lack of specific data required to follow the segmentation method described in the paper [3], this step has been carried out by applying the K-Means machine learning algorithm [8]. With this method, four different types of breast tissue have been identified: nodular, linear, homogeneous and radiolucent.

A. Method

Initial centroids: Initial centroids are set for both images to prevent unintended colour swaps during cluster formation.

Image flattening: The images are reshaped into one-dimensional arrays for clustering.

K-Means implementation: The K-Means algorithm is applied to the flattened images using the calculated initial centroids. The algorithm assigns each pixel to one of the clusters.

Label mapping: The clustered labels are reshaped back to the original dimensions of the images.

Colour mapping: The clustered image is coloured using a personalized colour map, with different colours for each cluster and black for the background. The colours are assigned as follows: for red, the nodular tissue; for gray, the radiolucent tissue; for blue, the linear tissue; and for yellow, the homogeneous tissue.

B. Visualization of results

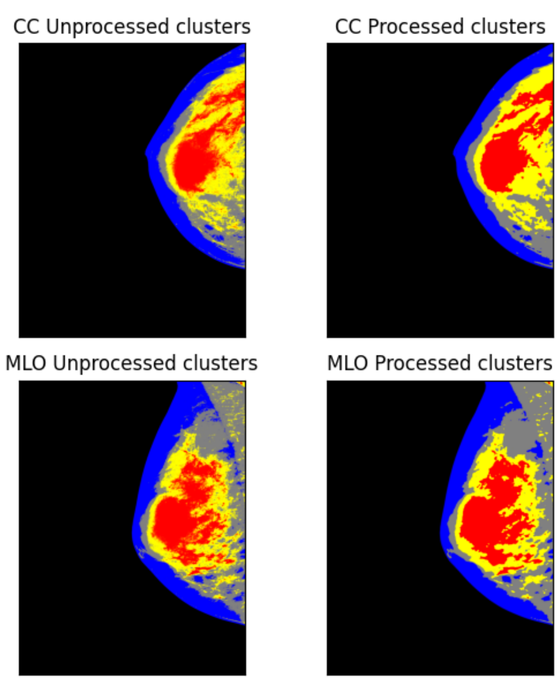

Fig. 6. The images on the left depict the unprocessed clustered images, while the images on the right depict the processed clustered images.

As shown in Fig. 6, the edges of the breast tissue of the processed images appear much more defined and clear, compared to the unprocessed ones. In addition to this, the amount of scattered points is reduced, the cluster regions of the processed images are much more defined and an increased contrast between regions is shown. This decreases the overall noise and makes it easier to distinguish the different tissues of the breast.

VII. Conclusions

In this project, we have proposed a comprehensive suite of pre-processing techniques to improve tissue segmentation of mammographic images. The approach that has been taken can handle a variety of problems that may arise in the process of mammographic image analysis, like uneven illumination and intensity variations.

The application of methods like periphery separation, intensity ratio propagation and breast thickness estimation has proven to not only make an image more uniform but also segment breast tissues in a much more clear manner. This can facilitate more accurate clinical analyses and interpretations, leading to an early detection and diagnosis of breast cancer.

To sum up, our results indicate that all of these preprocessing techniques considerably improve breast tissue visibility and segmentation. This work shows the real power of advanced image processing methods in clinical diagnosis.

References

[1] Dumitrescu, R. G., & Cotarla, I. (2007). *Understanding breast cancer risk – where do we stand in 2005?* Journal of Cellular and Molecular Medicine, 9(1), 208-221. [Online]. Available: https://doi.org/10.1111/j.1582-4934.2005.tb00350.x

[2] Wang, L. (2017). *Early Diagnosis of Breast Cancer*. Sensors, 17(7), 1572. [Online]. Available: https://doi.org/10.3390/s17071572

[3] He, W., Hogg, P., Juette, A., Denton, E. R. E., & Zwiggelaar, R. (2015). *Breast image pre-processing for mammographic tissue segmentation*. Computers in Biology and Medicine, 67, 61-73. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2015.10.002

[4] Ngx, T. (2012). *INbreast: toward a full-field digital mammographic database*. Kaggle. [Online]. Available: https://www.kaggle.com/datasets/tommyngx/inbreast2012

[5] Baeldung. (2023). *Otsu Segmentation*. Baeldung. [Online]. Available: https://www.baeldung.com/cs/otsu-segmentation

[6] Scikit-Image. (n.d.). *scikit-image API: skimage.morphology*. Scikit-Image Documentation. [Online]. Available: https://scikit-image.org/docs/stable/api/skimage.morphology.html

[7] Scikit-Image. (n.d.). *skimage.measure*. Scikit-Image Documentation. [Online]. Available: https://scikit-image.org/docs/stable/api/skimage.measure.html

[8] Scikit-Learn. (n.d.). *sklearn.cluster.KMeans*. Scikit-Learn Documentation. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html